DOCUMENT RESUME

ED 399 298 TM 025 613

AUTHOR Johanson, George A.; Doston, Glenn
TITLE Single-Item Measurement: Would You Recommend It to a Friend?
PUB DATE Apr 94
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Graduates; *Data Analysis; *Graduate Surveys; Higher Education; Likert Scales; *Program Evaluation; *Questionnaires; Surveys; *Test Construction; Test Items
IDENTIFIERS *Dichotomous Responses

ABSTRACT

Analyses of questionnaire data from a program evaluation indicate that the two dichotomous items "Would you recommend this to a friend?" and "Would you choose to do this again?" are not as interchangeable as might be expected from the survey literature. As part of the evaluation of a university program, a survey of graduates was conducted with a mailed questionnaire. The number of individual Likert-scored items ranged from 3 to 13 for each of the 9 subscales of interest. There were 59 usable responses to the survey, a response rate of 40%, but only 42 responses with complete data on all variables of interest. Responses to the recommend-to-a-friend option were more highly associated with the nine subscales related to program evaluation, and somewhat less influenced by personal situations of respondents. For evaluation purposes, the phrasing "Would you recommend this to a friend?" is recommended over the alternative. (Contains one table, one figure, and five references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY
GEORGE JOHANSON

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

ED 399 298

SINGLE-ITEM MEASUREMENT:

WOULD YOU RECOMMEND IT TO A FRIEND?

A paper presented at the 1994 annual meeting
of the American Educational Research Association,
New Orleans, LA.

George A. Johanson & Glenn Doston
Ohio University

ABSTRACT

Analyses of questionnaire data from a program evaluation indicate that the two dichotomous items 1. 'Would you recommend this to a friend?' and 2. 'Would you choose to do this again?' are not as interchangeable as might be expected from the survey literature. For evalutaion purposes, the former phrasing is recommended over the later for several reasons.

TM 025013

BEST COPY AVAILABLE

SINGLE-ITEM MEASUREMENT:

WOULD YOU RECOMMEND IT TO A FRIEND?

In the survey and evaluation literature, it is rather common to see a single overall or summative item of the form 'Would you recommend this to a friend?' or 'Would you choose to do this again?'. A brief examination of a portion of this literature would indicate that these questions are often seen as sufficiently similar to be used almost interchangeably in practice.

In addition, overall questions of this sort are sometimes seen to be of great importance. For example, in their annual questionnaire, Consumer's Union asks about automobiles 'Concerning all factors (price, performance, reliability, comfort, enjoyment, etc.) would you buy this car if you had it to do all over again?' (Staff, 1993). Responses to this single item are then reported as the sole measure of owner satisfaction.

Might owner satisfaction have been somewhat different if Consumer's Union had asked 'Concerning all factors (price, performance, reliability, comfort, enjoyment, etc.) would you recommend this car to a friend?' That is, might one form of this often used question be more effective than the other for general usage?

SINGLE-ITEM MEASUREMENT

In discussing the difficulty of single-item measurement of attitudes or values in general, Anderson et al. (1983) note that:

> The problem resides in our inability to get a response that measures only the relevant attitude. The response is usually a composite of various things such as item specific components, a general attitude component, and error of various sorts. Presumably, we are interested in the general attitude. If we have information only from one response per object per respondent, it is not possible to estimate or separate the components. Treating the response as a good measure (reliable and valid) is risky.
>
> It is well known that slight variations in question wording and format can produce large variations in response percentages. (p. 248)

Mueller (1986) states that reliability is the 'major reason' (p. 80) multiple items are used rather than single items. He goes on to note that single items may be adequately reliable for certain purposes (e.g., measuring group means), but not for others. Converse and Presser (1986) acknowledge that multiple

measures are the 'strategy of choice' (p. 45), largely because they may help the researcher understand the complexity of an attitude.

Given these caveats, if professional judgment is such that, even with the acknowledged limitations, a single overall item is needed, does it make any difference whether the item is 'recommend to a friend' or 'do again'?

THE SURVEY

As part of an evaluation of one of the programs at a midwestern university, a survey of graduates was conducted with a mailed questionnaire. There were 9 subscales of interest that related to: the objectives of the program (OBJECTIVES), the curriculum (CURRICULUM), the awareness of the students of opportunities in the field (AWARENESS), the program adequacy (PROGRAM), the quality of the instruction or teaching (TEACHING), the opportunity to interact with faculty (INTERACT), quality of faculty (FACULTY), the quality of fellow students (STUDENTS), and the opportunity to become involved in a variety of activities (INVOLVEMENT). The number of individual Likert (scored 1-5) items ranged from 3 to 13 for each of the subscales.

In addition, there were both forms of the overall item under consideration: 'WOULD YOU RECOMMEND THIS

PROGRAM TO A FRIEND?' (RFRIEND) and 'IF YOU HAD IT TO DO OVER, WOULD YOU ENTER THE PROGRAM AGAIN?' (DOAGAIN). There was a prompt to add comments if desired after each of these items.

In all, 148 surveys were mailed to all those graduates in the United States for whom a current location was available. There were 59 usable responses to the survey for a response rate of 40%.

RESULTS

In general, the program being evaluated was well liked with a majority of the respondents indicating that they would both recommend the program to a friend and would also enter the program again themselves. The 9 subscales were judged to have had satisfactory internal consistency reliabilities (Cronbach alpha ranged from .68 to .93).

Since there were only 42 subjects with complete data (listwise) on all of the variables of interest, a principal components analysis was performed to reduce the number of scales. Table 1 shows the correlations among the 9 subscales, DOAGAIN, and RFRIEND.

<insert Table 1 about here>

There were two factors (eigenvalues greater than one) which (cumulatively) explained 66.4% of the variation

in the 9 subscales. An oblique (oblimin) rotation indicated that the faculty-related subscales loaded largely on the first factor (54% of the variation) while the less faculty related subscales (STUDENTS, CURRICULUM, and OBJECTIVES) loaded more heavily on the second factor (12.4% of the variation). The correlation between the factors was .49. Factor scores were computed (FACTOR1 and FACTOR2, by regression methods) and the 9 subscales were judged to be adequately represented by these two correlated components and the corresponding factor scores.

The responses to RFRIEND correlated weakly with both FACTOR1 and FACTOR2 (-.28 for both, p=.07), but only FACTOR1 correlated significantly (-.32, p=.04) with DOAGAIN. The correlation of FACTOR2 with DOAGAIN was -.04, p=.78. Since DOAGAIN and RFRIEND were dichotomously scored, these are point-biserial correlations and are negative due to the scoring of DOAGAIN, RFRIEND (YES=1, NO=2), and the 9 subscales (higher scores meant more favorable opinions).

Using FACTOR1 and FACTOR2 as dependent variables, a two-way factorial multivariate analysis of variance was conducted with RFRIEND and DOAGAIN as independent variables each with two levels (YES, NO) each. Forty-two cases had complete data and were used in the MANOVA. As anticipated, there was a significant main

effect for RFRIEND (Wilks Lambda=.66, approximate $F=9.69_{(2,37)}$, p=.00), but not for DOAGAIN (Wilks Lambda=.91, approximate $F=1.88_{(2,37)}$, p=.17). The only significant univariate main effect ($F=5.21_{(1,38)}$, p=.03) was for FACTOR2 and for RFRIEND.

Somewhat surprising to us, the interaction of DOAGAIN and RFRIEND was significant both at the multivariate level (Wilks Lambda=.84, p=.04) and univariate level with FACTOR1 ($F=4.78_{(1,38)}$, p=.04), but not with FACTOR2 ($F=0.04_{(1,38)}$, p=.85). Figure 1 shows the group means for the first factor score.

<insert Figure 1 about here>

DISCUSSION

Due primarily to the low response rate and small sample size, our findings are tentative, but do indicate that the overall questions 'Would you do this again?' and 'Would you recommend this to a friend?' were far from interchangeable (although they were significantly related, phi=0.62) for these respondents. It would appear that the second phrasing of the item was superior for evaluation purposes since responses to this item were more highly associated with the 9 subscales relating to program evaluation. For example, 5 of the 9 listwise correlations were statistically

significant at p<.01 with RFRIEND; only 1 of the 9 with DOAGAIN at p<.01, see Table 1.

The probes following both RFRIEND and DOAGAIN yielded important supplementary information. One respondent who answered 'NO' to DOAGAIN noted that the program was 'the only option at the time'. While this was contradictory (if you would not take the only option, what is left?), it also indicated that unique personal circumstances might well induce a person to act in a manner that is not entirely consistent with their evaluation of the object under consideration. As another illustration, a second respondent who also answered 'NO' to DOAGAIN stated that he or she 'would choose a different program due to money and family issues', but commented on RFRIEND (where the response was also 'NO') that the 'class sizes were too large'. Note that the former response was seemingly based on more personal circumstances, while the latter response was more related to the program evaluation as intended.

Still another respondent indicated 'YES' to RFRIEND with the comment 'basically a good program', but 'NO' to DOAGAIN with the comment 'dangerous profession, unwarranted lawsuits'. The former was more related to program evaluation, the latter more related to perceived occupational hazards. As a final example, one respondent chose 'NO' for RFRIEND with the comment

that the 'program has reputation for quickest, easiest masters at university', but then chose 'YES' for DOAGAIN and commented that this was for the same reason given for RFRIEND 'plus supplemented w/ continuing education courses to preserve integrity'. The notion would appear to be that while the program was seen to be deficient, personal efforts may compensate for the deficiency. Clearly, the response most related to program evaluation for this respondent is that to RFRIEND.

The interaction of DOAGAIN and RFRIEND with respect to FACTOR1 (Figure 1) would seem to add support to the interpretations of the above comments when we note that the difference in program evaluation (as measured by FACTOR1) is greatest when the response to RFRIEND is 'YES' and DOAGAIN is 'NO'. Recall that FACTOR1 is the factor score associated with the largest eigenvalue of the 9 evaluation subscales and, as such, can be thought of as a summary measure of the evaluation. The idea that a person may have seen no other option at the time, had personal factors entering into the decision, or even possessed personal qualities that might compensate for the evaluation, would resolve the interaction with such a summary measure. Note, however, that when the responses to both DOAGAIN and RFRIEND are 'NO', the program evaluation is very

similar; the evaluation is highest when the responses to both questions are 'YES', as expected.

Still another factor that may play a role in differentiating between these items is the extent to which each is speculative or hypothetical. That is, while RFRIEND is typically rather realistic, DOAGAIN is sometimes quite imaginary (as in our survey). Recommendations for practice often indicate that such hypothetical questions are of very limited value. (Moser & Kalton, 1971, p. 326).

Since this was not a designed experiment, we were not able to control the location of RFRIEND and DOAGAIN (they followed the other evaluation items) nor the order of the items themselves (DOAGAIN immediately followed RFRIEND). Also, the phrasing of neither DOAGAIN nor RFRIEND was varied. Therefore, in addition to limited generalizability, we must add concerns about phrasing, location and order. Finally, the responses to DOAGAIN and RFRIEND were dichotomous; the correlations, at least, would likely be different were these on, say, a five- or seven-point scale. As we indicated earlier, our findings are tentative or suggestive.

In conclusion, we might reiterate the Anderson et al. (1983) caveat that single-item measurement is difficult because evaluation is 'usually a composite of

various things'. In the case of the items 'recommend to a friend' and 'would do again', it would appear that the later phrasing is more susceptible to non-evaluative influences such as personal circumstances at the time of the decision and, in addition, may be quite hypothetical. The 'would do again' item may thus be less useful as an overall or summative item for evaluation purposes. The 'recommend to a friend' phrasing would appear to be less speculative and less influenced by personal considerations and hence a better choice for many purposes.

REFERENCES

Anderson, A. B., Basilevsky, A., & Hum, D. P. J. 1983. Measurement: Theory and techniques. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), Handbook of survey research (pp. 231-287). New York: Academic Press.

Converse, J. M., & Presser, S. 1986. Survey questions: Handcrafting the standardized questionnaire. Beverly Hills, CA: Sage Publications.

Moser, C. A., & Kalton, G. 1971. Survey methods in social investigation (2nd ed.). London: Heinemann.

Mueller, D. J. 1986. Measuring social attitudes: A handbook for researchers and practitioners. New York, NY: Teacher's College Press.

Staff. 1993, April. "Would you buy that car again?" Consumer Reports, 256.

Table 1

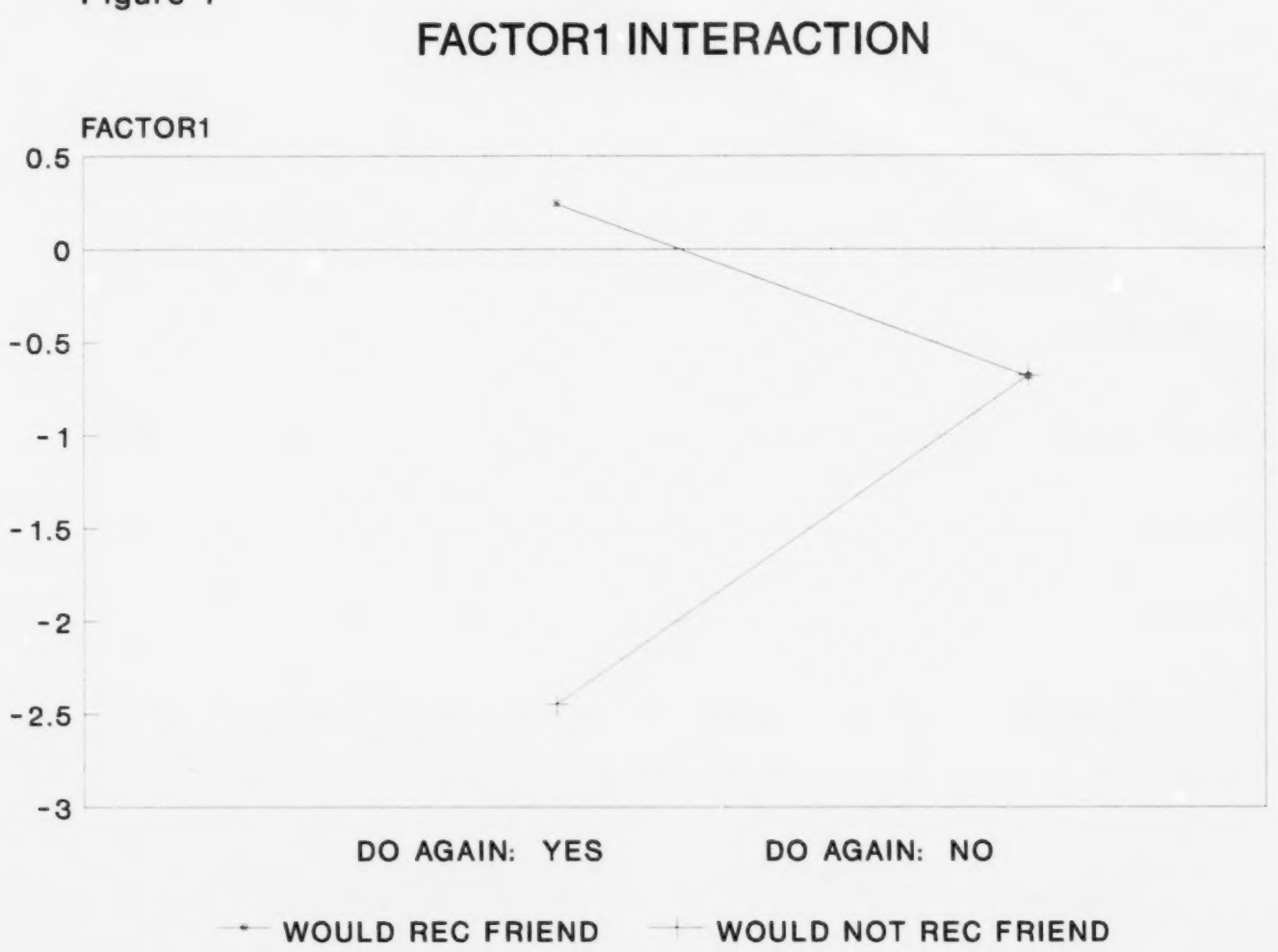
Correlations Between DOAGAIN, RFRIEND, and the (Nine) Evaluation Subscales

Subscale	RFRIEND	DOAGAIN	OBJECTV	CURRICL	AWARENS	PROGRAM	TEACHNG	INTRACT	FACULTY	STUDENT	INVOLVE
RFRIEND	1.0000										
DOAGAIN	.6159**	1.0000									
OBJECTV	-.5010**	-.1533	1.0000								
CURRICL	-.4251**	-.1268	.7115**	1.0000							
AWARENS	-.1955	-.0826	.3058*	.4314**	1.0000						
PROGRAM	-.4087**	-.1930	.3711*	.6047**	.4828**	1.0000					
TEACHNG	-.3558*	-.1536	.4496**	.6409**	.5426**	.6326**	1.0000				
INTRACT	-.3931*	-.3725*	.2929	.4575**	.6458**	.5612**	.5856**	1.0000			
FACULTY	-.6438**	-.4607**	.4777**	.6357**	.5536**	.6377**	.6304**	.7073**	1.0000		
STUDENT	-.1739	-.0185	.3353*	.4557**	.1984	.2677	.3805*	.1105	.3420*	1.0000	
INVOLVE	-.4935**	-.3441*	.3765*	.3806*	.5213**	.3018	.4382**	.6086**	.7469**	.3701*	1.0000

Notes. * = Significance less than or equal .05 (2-tailed), ** = Significance less than or equal .01 (2-tailed), N=42 (listwise).

BEST COPY AVAILABLE

Figure 1


FACTOR1 INTERACTION
FACTOR1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
DO AGAIN: YES
DO AGAIN: NO
WOULD REC FRIEND
WOULD NOT REC FRIEND